

---

# Bayesian Predictive Power: Theory, challenges in implementation, and perspectives

*Kaspar Rufibach*

*Methods, Collaboration & Outreach Group, Department of Biostatistics, Roche Basel  
Bayes-pharma Lyon, 23rd May 2019*

Generated 2019-05-21 at 15:27:17.



# Acknowledgments

Co-authors: Markus Abt, Hans Ulrich Burger, Paul Jordan.

# Acknowledgments

Co-authors: Markus Abt, Hans Ulrich Burger, Paul Jordan.

Not meant to be a comprehensive introduction to the topic.

# Probability of Success

**Comprehensive** assessment of probability to meet “target” of Phase 3 trial. Based on:

- 1 **Quantitative** assessment: Bayesian predictive power,
- 2 **Qualitative** adjustment: non-quantifiable additional information e.g. on competitors, uncertainty around assumptions, change in endpoint from Phase 2 to Phase 3, safety, ...

“Target”: be significant or beat “target product profile” (TPP).

Used for:

- Calculating project **valuations**,
- supporting **funding**, trade-off, and gating decisions by senior management,
- developing **budgets** and **hiring** plans,
- planning manufacturing **capacity**,
- allocate resources in **plansource** for not-yet approved molecules.

**How can we quantify  
probability of success?**

# Bayesian Predictive Power

Continuous endpoint, true effect  $\Delta$ , estimator assumed to follow Normal distribution.

Estimate  $\hat{\Delta}_{\text{final}}$  at final analysis of pivotal trial, based on  $n_{\text{final}}$  observations:

$$\hat{\Delta}_{\text{final}} \sim N(\Delta, \sigma_{\text{final}}^2 = \sigma^2 / n_{\text{final}}).$$

Pivotal trial is called a success if  $\hat{\Delta}_{\text{final}} \leq \Delta_{\text{suc}}$  (think of log hazard ratio).

$\Delta_{\text{suc}}$ : can be

- **Minimal detectable difference** (MDD), i.e. effect size such that trial is “just significant”.
- Any **other quantity of interest**, e.g. alternative that gives 80% power  $\Rightarrow$  target product profile (TPP).

# Bayesian Predictive Power

Quantity of interest = **power function**:

$$P(\hat{\Delta}_{\text{final}} \leq \Delta_{\text{suc}}) = \Phi\left(\frac{\Delta_{\text{suc}} - \Delta}{\sigma_{\text{final}}}\right).$$

Depends on true effect  $\Delta \Rightarrow$  assume distribution over  $\Delta$  with density  $q$  and average:

$$\begin{aligned}\text{PoS} &= \mathbb{E}_{\Delta}\left(P_{\Delta}(\hat{\Delta}_{\text{final}} \leq \Delta_{\text{suc}})\right) \\ &= \int_{-\infty}^{\infty} \Phi\left(\frac{\Delta_{\text{suc}} - \Delta}{\sigma_{\text{final}}}\right) q(\Delta) d\Delta.\end{aligned}$$

**Bayesian predictive power** initially introduced in Spiegelhalter et al. (1986).

# Bayesian Predictive Power

Terminology:

- Assurance O'Hagan et al. (2005),
- average power,
- hybrid classical-Bayesian,
- probability of Success (PoS),
- ...

We use **PoS**.



# Quantities

- 1 **Power**  $P_{\Delta}(\hat{\Delta}_{\text{final}} \leq \Delta_{\text{suc}})$ . At trial start, function of assumed effect  $\Delta$ .
- 2 **Conditional power**:  $P_{\Delta}(\hat{\Delta}_{\text{final}} \leq \Delta_{\text{suc}} | \hat{\Delta}_{\text{interim}} = \Delta_{\text{interim}})$ . “Updated” power after trial has started, function of  $\Delta$  and  $\Delta_{\text{interim}}$ .
- 3 **Bayesian predictive power** (BPP): average over (conditional) power with respect to distribution over  $\Delta$ .
- 4 **Predictive probability**:  $\mathbb{E}(\text{posterior of clinically meaningful effect} \mid \text{every possible future outcome})$ . See e.g. [Berry et al. \(2011\)](#).

Different quantities that

- depend on different assumptions,
- have different properties,
- have different interpretations.

# Time-to-event endpoint

Approximate distribution of **estimated log(hazard ratio)**  $\hat{\theta} := \log \hat{\lambda}$ :

$$\hat{\theta} \approx N(\theta, 4/d).$$

- $\theta = \log \lambda$ : **true underlying effect**, true log-hazard ratio.
- 1:1 randomized trial:  $\text{Var}(\hat{\theta}) = 4/d$ .
- $d$ : total number of events in both arms.

In context of pivotal trial:

- Random variable  $\hat{\theta}_{\text{final}} \sim N(\theta, \sigma_{\text{final}}^2 = 4/d_{\text{final}})$ .
- $d_{\text{final}}$ : number of events at final analysis.
- $\alpha_{\text{final}}$ : significance level at final analysis. May be adjusted for group-sequential design.

# Example

Assumptions:

- Phase 2 result:  $\hat{\theta}_{\text{Phase 2}} = \log(0.700)$ , based on  $d_{\text{prior}} = 50$  events.
  - Consider prior fixed, do not account for uncertainty in estimation of parameters.
- Phase 3: 80% power to detect hazard ratio 0.74.
- Final analysis after  $d_{\text{final}} = 352$  events.
- $\alpha_{\text{final}} = 0.046$  (efficacy interim after 67% of information).
- Minimal detectable difference at final analysis:  $\theta_{\text{suc}} = \log(0.809)$ .

PoS at start of Phase 3, assuming we know Phase 2 result:

$$\text{PoS} = \int_{-\infty}^{\infty} P_{\theta}(\hat{\theta}_{\text{final}} \leq \theta_{\text{suc}}) \phi_{\mu=\log(0.700), \sigma^2=4/50}(\theta) d\theta = 0.683.$$

**How can we update PoS after  
not stopping at an interim analysis?**

# What if the sponsor remains blinded to $\hat{\theta}_{\text{interim}}$ ?

Assume:

- **Binding** interim analysis for **futility** and **efficacy**: continue trial only if estimated log hazard ratio  $\hat{\theta}_{\text{interim}} \in \mathcal{I}_{\text{interim}} := [\theta_{\text{efficacy}}, \theta_{\text{futility}}]$ .
- Interim assessment typically done by independent data monitoring committee (iDMC). Sponsor remains **blinded**: only informed whether trial is stopped or continued.

Trial **not** stopped at this interim  $\Rightarrow$  sponsor knows  $\hat{\theta}_{\text{interim}} \in \mathcal{I}_{\text{interim}}$ .

$\theta_{\text{efficacy}} = -\infty$  or  $\theta_{\text{futility}} = \infty$  easily possible.

Knowledge **internal** to the trial.

How can we update PoS after such an interim?

# Update PoS after blinded interim

PoS formula...

$$\text{PoS} = \int_{-\infty}^{\infty} P_{\theta}(\hat{\theta}_{\text{final}} \leq \theta_{\text{suc}}) q_{\text{prior}}(\theta) d\theta$$

# Update PoS after blinded interim

PoS formula...

$$\text{PoS} = \int_{-\infty}^{\infty} P_{\theta}(\hat{\theta}_{\text{final}} \leq \theta_{\text{suc}}) q_{\text{prior}}(\theta) d\theta$$

...becomes:

$$\text{PoS} = \int_{-\infty}^{\infty} P_{\theta}(\hat{\theta}_{\text{final}} \leq \theta_{\text{suc}} | \hat{\theta}_{\text{interim}} \in \mathcal{I}_{\text{interim}}) q_{\text{posterior}}(\theta) d\theta.$$

## Computation of $P_{\theta}(\hat{\theta}_{\text{final}} \leq \theta_{\text{suc}} | \hat{\theta}_{\text{interim}} \in \mathcal{I}_{\text{interim}})$

Computation:  $\hat{\theta}_{\text{interim}}$  and  $\hat{\theta}_{\text{final}}$  are **correlated**!



## Computation of $P_{\theta}(\hat{\theta}_{\text{final}} \leq \theta_{\text{suc}} | \hat{\theta}_{\text{interim}} \in \mathcal{I}_{\text{interim}})$

Computation:  $\hat{\theta}_{\text{interim}}$  and  $\hat{\theta}_{\text{final}}$  are **correlated**!

Apply **canonical joint distribution** for group-sequential tests:

$$\begin{pmatrix} \hat{\theta}_{\text{interim}} \\ \hat{\theta}_{\text{final}} \end{pmatrix} \sim N\left(\begin{pmatrix} \theta \\ \theta \end{pmatrix}, \begin{pmatrix} 4/d_{\text{interim}} & 4/d_{\text{final}} \\ 4/d_{\text{final}} & 4/d_{\text{final}} \end{pmatrix}\right).$$

## Computation of $P_{\theta}(\hat{\theta}_{\text{final}} \leq \theta_{\text{suc}} | \hat{\theta}_{\text{interim}} \in \mathcal{I}_{\text{interim}})$

Computation:  $\hat{\theta}_{\text{interim}}$  and  $\hat{\theta}_{\text{final}}$  are **correlated**!

Apply **canonical joint distribution** for group-sequential tests:

$$\begin{pmatrix} \hat{\theta}_{\text{interim}} \\ \hat{\theta}_{\text{final}} \end{pmatrix} \sim N\left(\begin{pmatrix} \theta \\ \theta \end{pmatrix}, \begin{pmatrix} 4/d_{\text{interim}} & 4/d_{\text{final}} \\ 4/d_{\text{final}} & 4/d_{\text{final}} \end{pmatrix}\right).$$

**Conditional probability:**

$$\begin{aligned} P_{\theta}(\hat{\theta}_{\text{final}} \leq \theta_{\text{suc}} | \hat{\theta}_{\text{interim}} \in \mathcal{I}_{\text{interim}}) &= \\ &= \frac{P_{\theta}(\hat{\theta}_{\text{final}} \leq \theta_{\text{suc}}, \hat{\theta}_{\text{interim}} \in \mathcal{I}_{\text{interim}})}{P_{\theta}(\hat{\theta}_{\text{interim}} \in \mathcal{I}_{\text{interim}})} \\ &= \frac{P_{\theta}\left(\begin{pmatrix} \theta_{\text{efficacy}} \\ -\infty \end{pmatrix} \leq \begin{pmatrix} \hat{\theta}_{\text{interim}} \\ \hat{\theta}_{\text{final}} \end{pmatrix} \leq \begin{pmatrix} \theta_{\text{futility}} \\ \theta_{\text{suc}} \end{pmatrix}\right)}{\Phi\left((\theta_{\text{futility}} - \theta)/\sqrt{4/d_{\text{interim}}}\right) - \Phi\left((\theta_{\text{efficacy}} - \theta)/\sqrt{4/d_{\text{interim}}}\right)}. \end{aligned}$$

## Computation of $P_{\theta}(\hat{\theta}_{\text{final}} \leq \theta_{\text{suc}} | \hat{\theta}_{\text{interim}} \in \mathcal{I}_{\text{interim}})$

Computation:  $\hat{\theta}_{\text{interim}}$  and  $\hat{\theta}_{\text{final}}$  are **correlated**!

Apply **canonical joint distribution** for group-sequential tests:

$$\begin{pmatrix} \hat{\theta}_{\text{interim}} \\ \hat{\theta}_{\text{final}} \end{pmatrix} \sim N\left(\begin{pmatrix} \theta \\ \theta \end{pmatrix}, \begin{pmatrix} 4/d_{\text{interim}} & 4/d_{\text{final}} \\ 4/d_{\text{final}} & 4/d_{\text{final}} \end{pmatrix}\right).$$

**Conditional probability:**

$$\begin{aligned} P_{\theta}(\hat{\theta}_{\text{final}} \leq \theta_{\text{suc}} | \hat{\theta}_{\text{interim}} \in \mathcal{I}_{\text{interim}}) &= \\ &= \frac{P_{\theta}(\hat{\theta}_{\text{final}} \leq \theta_{\text{suc}}, \hat{\theta}_{\text{interim}} \in \mathcal{I}_{\text{interim}})}{P_{\theta}(\hat{\theta}_{\text{interim}} \in \mathcal{I}_{\text{interim}})} \\ &= \frac{P_{\theta}\left(\begin{pmatrix} \theta_{\text{efficacy}} \\ -\infty \end{pmatrix} \leq \begin{pmatrix} \hat{\theta}_{\text{interim}} \\ \hat{\theta}_{\text{final}} \end{pmatrix} \leq \begin{pmatrix} \theta_{\text{futility}} \\ \theta_{\text{suc}} \end{pmatrix}\right)}{\Phi\left((\theta_{\text{futility}} - \theta)/\sqrt{4/d_{\text{interim}}}\right) - \Phi\left((\theta_{\text{efficacy}} - \theta)/\sqrt{4/d_{\text{interim}}}\right)}. \end{aligned}$$

**mvtnorm** in R.

# Computation of $q_{\text{posterior}}$

In principle,  $q_{\text{prior}}$  quantifies update in **external** knowledge.

Sensible to adjust for knowledge " $\hat{\theta}_{\text{interim}} \in \mathcal{I}_{\text{interim}}$ ".

From **Bayes'** theorem:

$$\begin{aligned} q_{\text{posterior}}(\theta | \hat{\theta}_{\text{interim}} \in \mathcal{I}_{\text{interim}}) &\propto P(\hat{\theta}_{\text{interim}} \in \mathcal{I}_{\text{interim}} | \theta) q_{\text{prior}}(\theta) \\ &= \left( \Phi\left(\frac{\theta_{\text{futility}} - \theta}{\sqrt{4/d_{\text{interim}}}}\right) - \Phi\left(\frac{\theta_{\text{efficacy}} - \theta}{\sqrt{4/d_{\text{interim}}}}\right) \right) q_{\text{prior}}(\theta). \end{aligned}$$

# Interim analysis

PoS at start of Phase 3, assuming we know Phase 2 result:

$$\text{PoS} = \int_{-\infty}^{\infty} P_{\theta}(\hat{\theta}_{\text{final}} \leq \theta_{\text{suc}}) \phi_{\mu=\log(-0.357), \sigma^2=4/50}(\theta) d\theta = \mathbf{0.683}.$$

Interim after 67% of information = 236 events.

**Efficacy** interim:

- Significance level  $\alpha_{\text{interim}} = 0.012$ .
- Minimal detectable difference:  $\theta_{\text{suc}}^{\text{interim}} = 0.722$ .

**Futility** interim: continue if  $\text{HR} \leq 1$ .

**How does PoS change if we  
do not stop at a futility interim?**

# Futility interim analysis only

Futility interim passed with boundary  $HR \leq 1$ : we know that

- $0 < HR \leq 1$  or
- $\hat{\theta}_{\text{interim}} \in (-\infty, \log(1)]$ .

How does PoS change after this interim?

# Futility interim analysis only

Futility interim passed with boundary  $HR \leq 1$ : we know that

- $0 < HR \leq 1$  or
- $\hat{\theta}_{\text{interim}} \in (-\infty, \log(1)]$ .

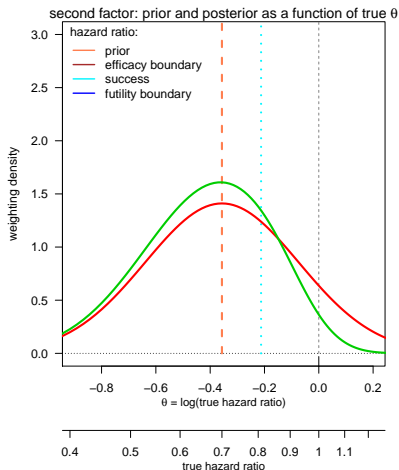
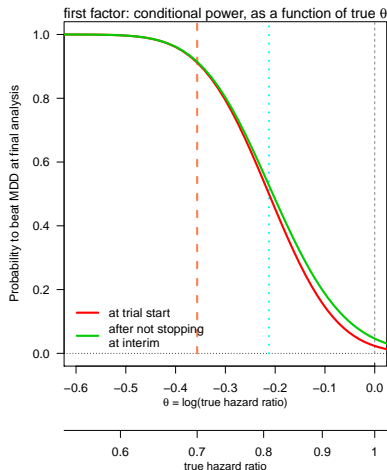
How does PoS change after this interim?

PoS increases from **0.683** to **0.782**.



Why?

# Futility interim analysis only - plot both factors in PoS formula



Green density **not** a Normal density, but product of difference of values of Normal CDF and prior density  $q_{\text{prior}}$ .

# Futility interim analysis only - sensitivity analysis

PoS at trial start: 0.683.

In our example, with  $\theta_{\text{efficacy}} = \log(0)$  and  $\theta_{\text{futility}} = \log(1)$ :

$\hat{\theta}_{\text{interim}}$	value of $\hat{\theta}_{\text{interim}}$	PoS after not stopping
$\in (\theta_{\text{efficacy}}, \theta_{\text{futility}}]$	$(-\infty, \log(1)]$	0.782
	$\log(0.5)$	1.000
$= \theta_{\text{futility}}$	$\log(1)$	0.004

# Futility interim analysis only - comments

After not stopping at interim, PoS increases from **0.683** to **0.782**.

Why does PoS **increase** after not stopping?

- Prior with prior mean  $\log(0.7)$  assigns weight to hazard ratios smaller than hazard ratio to finally beat,  $\theta_{\text{suc}} = \log(0.809)$ .
- **Not stopping** shifts mass of prior  $q_{\text{prior}}$  to the left of 1 for  $q_{\text{posterior}} \Rightarrow$  more weight on hazard ratios  $\leq \theta_{\text{suc}}$ .
- Together with small increase in conditional power accounts for **higher PoS** after not stopping.

# Does PoS decrease or increase after not stopping?

Depends on configuration of

- Prior mean  $\theta_0 = \log(0.700)$ ,
- efficacy interim boundary  $\theta_{\text{efficacy}} = \log(0.722)$ ,
- minimal detectable difference at final analysis  $\theta_{\text{suc}} = \log(0.809)$ ,
- futility interim boundary  $\theta_{\text{futility}} = \log(1.000)$ .



## Sequentially updating the likelihood of success of a Phase 3 pivotal time-to-event trial based on interim analyses or external information

Kaspar Rufibach, Paul Jordan, and Markus Abt

F. Hoffmann-La Roche Ltd., Product Development Biostatistics, Basel, Switzerland

### ABSTRACT

When performing a pivotal clinical trial, it may be of interest to assess the probability of success (PoS) of that trial. Initially evaluated when the trial is designed, PoS can be updated as the trial progresses and new information about the drug effect becomes available. Such information can be external to the trial, such as results from trials conducted in parallel, or internal, such as continuing after an interim analysis. We develop a framework to update PoS based on such internal and external information for a time-to-event endpoint and illustrate it using a realistic development program for a new molecule.

### ARTICLE HISTORY

Received 8 May 2014

Accepted 30 September 2014

### KEYWORDS

Bayesian predictive power; conditional power; interim analysis; prior distribution; probability of technical success

Rule-of-thumb: initial PoS 40%, pass futility interim after 1/3 of events: PoS  $\approx$  60%, do not stop at efficacy interim after 2/3 of events: PoS  $\approx$  30%.

**What about  
choice of prior?**

**And what about the  
bathtub effect?**



# Choice of prior

So far **Normal prior**.

Flat prior often associated with **non-informativeness**.

Not necessarily the case for PoS!

See [Rufibach et al. \(2016a\)](#) for details.

What is the problem?

# Recall definitions and example

Power function:

$$T(\theta) \quad := \quad P_{\theta}(\hat{\theta}_{\text{final}} \leq \theta_{\text{suc}}) = \Phi\left(\frac{\theta_{\text{suc}} - \theta}{\sigma_{\text{final}}}\right).$$

PoS is expected power:

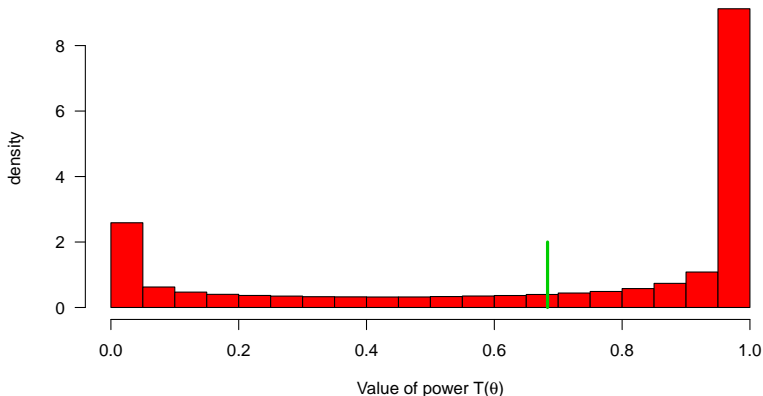
$$\text{PoS} = \mathbb{E}_{\theta} T(\theta) = \int_{-\infty}^{\infty} P_{\theta}(\hat{\theta}_{\text{final}} \leq \theta_{\text{suc}}) q_{\text{prior}}(\theta) d\theta.$$

Compute PoS via simulation (law of large numbers):

- Draw a sample  $(\hat{\theta}_1, \dots, \hat{\theta}_M)$  from prior.
- Compute  $T(\hat{\theta}_1), \dots, T(\hat{\theta}_M)$ .
- PoS = average over these values.

# Simulate PoS in example

Histogram of values of  $T(\theta)$  for  $\theta$  sampled from Normal prior  
sample size: 1'000'000



- ④ Is mean really appropriate number to **summarize** this histogram?
- ② Can we compute this density?

## Density of power $T(\Theta)$

Assume prior r.v.  $\Theta$  with PDF  $q$ , CDF  $Q$ , and define  $Y := T(\Theta)$  with PDF  $g$ , CDF  $G$ .

Use **transformation theorem** and **rule about derivative of an inverse** to get:

$$\begin{aligned}G(y) &= 1 - Q(\theta_{\text{suc}} - \sigma_{\text{final}}z), \\g(y) &= q(\theta_{\text{suc}} - \sigma_{\text{final}}z) \frac{\sigma_{\text{final}}}{\phi(z)}\end{aligned}$$

with  $z := \Phi^{-1}(y)$  and  $\phi$  the standard Normal density function.

For Normal prior  $\Theta \sim N(\theta_0, \sigma_0^2)$ :

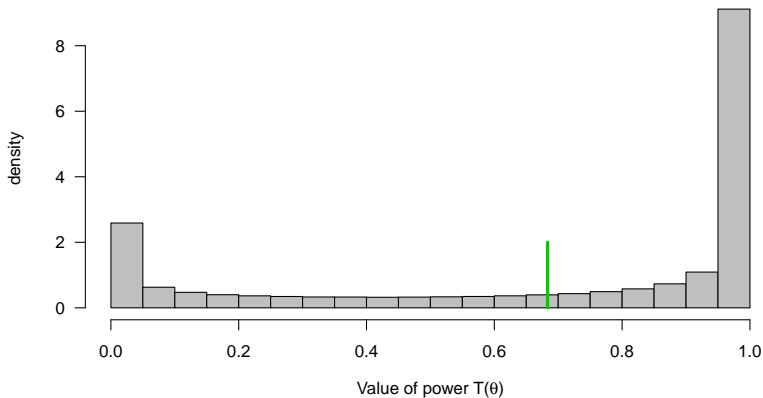
$$\begin{aligned}G(y) &= 1 - \Phi(\beta - \alpha z), \\g(y) &= \alpha \phi(\beta - \alpha z) [\phi(z)]^{-1},\end{aligned}$$

with

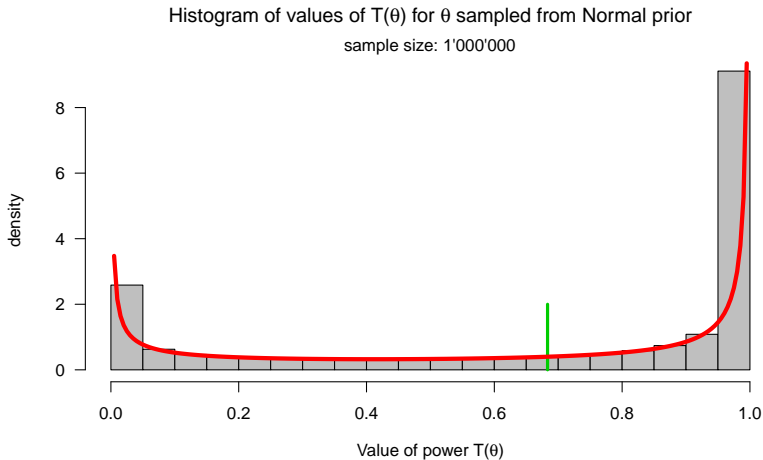
$$\begin{aligned}\alpha &= \sigma_{\text{final}}/\sigma_0 > 0, \\ \beta &= (\theta_{\text{suc}} - \theta_0)/\sigma_0.\end{aligned}$$

# Simulate PoS in example

Histogram of values of  $T(\theta)$  for  $\theta$  sampled from Normal prior  
sample size: 1'000'000

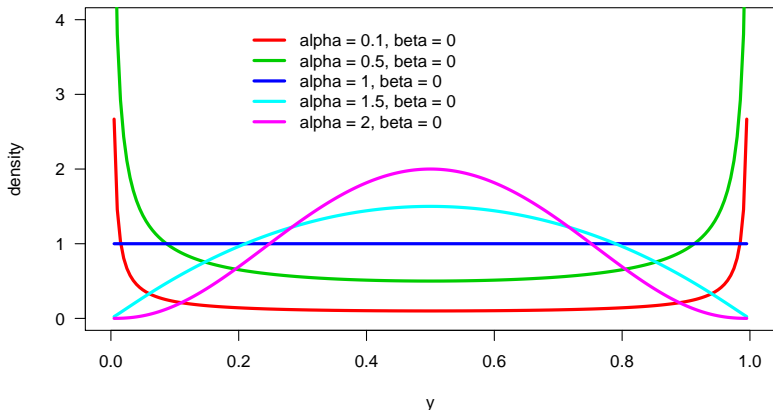


# Simulate PoS in example



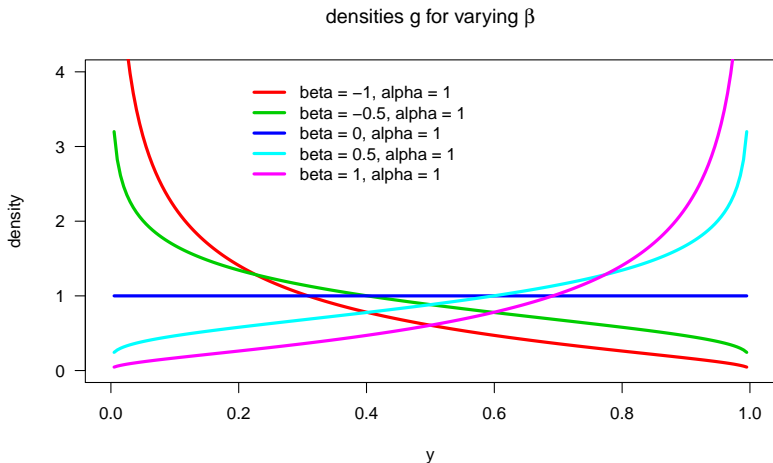
# Density $g$ as a function of $\alpha$ , for $\beta = 0$

densities  $g(y)$  for varying  $\alpha$



- When summarizing  $g$  with PoS  $\Rightarrow$  unimodal density most sensible?
- $\alpha = 1$ : transition between “bathtub-shaped” (even convex?) and unimodal (obviously not concave).
- Make qualitative features precise.

# Density $g$ as a function of $\beta$ , for $\alpha = 1$



- $\beta$  determines skewness of  $g$ .
- Make qualitative features precise.



# Qualitative features of $g$

## Theorem (Qualitative features of $g$ )

We have the following statements:

① If  $\alpha = 1$ , then  $g$  is

$$\begin{cases} \text{strictly decreasing for} & \beta < 0, \\ \text{constant for} & \beta = 0, \\ \text{strictly increasing for} & \beta > 0. \end{cases}$$

on  $[0, 1]$ . Minima and maxima of  $g$  are accordingly either at 0 or 1.

② If  $\alpha \neq 1$  then  $g$

$$\begin{cases} \text{has a minimum at } y_m \text{ if} & \alpha < 1, \\ \text{has a maximum at } y_m \text{ if} & \alpha > 1, \end{cases}$$

for  $y_m = \Phi(\alpha\beta/(\alpha^2 - 1))$ . Furthermore,  $g$

$$\begin{cases} \text{is decreasing for } y < y_m \text{ and increasing for } y > y_m \text{ if} & \alpha < 1, \\ \text{is increasing for } y < y_m \text{ and decreasing for } y > y_m \text{ if} & \alpha > 1. \end{cases}$$

**Proof:** Compute  $g', g''$ , discuss these.

## Why? And what does it mean?

Simplest case:  $\alpha = \beta = 0 \Rightarrow d_{\text{prior}} = d_{\text{final}}, \theta_0 = \theta_{\text{suc}} \Rightarrow g$  uniform.

# Why? And what does it mean?

Simplest case:  $\alpha = \beta = 0 \Rightarrow d_{\text{prior}} = d_{\text{final}}, \theta_0 = \theta_{\text{suc}} \Rightarrow g$  uniform.

Prior and distribution of pivotal effect size have same variance  $\Rightarrow$  power becomes uniform, either you beat  $\theta_{\text{suc}}$  with  $\hat{\theta}_{\text{final}}$  or not, with equal probability.

# Why? And what does it mean?

Simplest case:  $\alpha = \beta = 0 \Rightarrow d_{\text{prior}} = d_{\text{final}}, \theta_0 = \theta_{\text{suc}} \Rightarrow g$  uniform.

Prior and distribution of pivotal effect size have same variance  $\Rightarrow$  power becomes uniform, either you beat  $\theta_{\text{suc}}$  with  $\hat{\theta}_{\text{final}}$  or not, with equal probability.

**Why P(extreme PoS values) so high if  $\alpha < 1$ ?**  $d_0 < d_{\text{final}} \Rightarrow$  high variance of prior  $\Rightarrow$  high probability to have extreme HRs  $\Rightarrow$  power for these is either almost 0 or 1.

# Why? And what does it mean?

Simplest case:  $\alpha = \beta = 0 \Rightarrow d_{\text{prior}} = d_{\text{final}}, \theta_0 = \theta_{\text{suc}} \Rightarrow g$  uniform.

Prior and distribution of pivotal effect size have same variance  $\Rightarrow$  power becomes uniform, either you beat  $\theta_{\text{suc}}$  with  $\hat{\theta}_{\text{final}}$  or not, with equal probability.

**Why P(extreme PoS values) so high if  $\alpha < 1$ ?**  $d_0 < d_{\text{final}} \Rightarrow$  high variance of prior  $\Rightarrow$  high probability to have extreme HRs  $\Rightarrow$  power for these is either almost 0 or 1.

$g$  unimodal if  $\alpha > 1 \Rightarrow \sigma_{\text{final}} > \sigma_0 \Rightarrow d_{\text{final}} < d_0 \Rightarrow$  prior number of events **larger** than Phase 3 events.

# Why? And what does it mean?

Simplest case:  $\alpha = \beta = 0 \Rightarrow d_{\text{prior}} = d_{\text{final}}, \theta_0 = \theta_{\text{suc}} \Rightarrow g$  uniform.

Prior and distribution of pivotal effect size have same variance  $\Rightarrow$  power becomes uniform, either you beat  $\theta_{\text{suc}}$  with  $\hat{\theta}_{\text{final}}$  or not, with equal probability.

**Why P(extreme PoS values) so high if  $\alpha < 1$ ?**  $d_0 < d_{\text{final}} \Rightarrow$  high variance of prior  $\Rightarrow$  high probability to have extreme HRs  $\Rightarrow$  power for these is either almost 0 or 1.

$g$  unimodal if  $\alpha > 1 \Rightarrow \sigma_{\text{final}} > \sigma_0 \Rightarrow d_{\text{final}} < d_0 \Rightarrow$  prior number of events **larger** than Phase 3 events.

**Unrealistic** in clinical development.

# Priors explored

How should we choose prior to get **unimodal** PoS distribution?

Explored priors:

- truncated Normal,
- Uniform,
- Uniform prior with Normal tails.

None of them provides a unimodal density of power values under realistic assumptions.

Prior potentially informs BPP substantially.

Rufibach et al. (2016a).

**PoS is always smaller  
than power?!**



**PoS is always smaller  
than power?!**

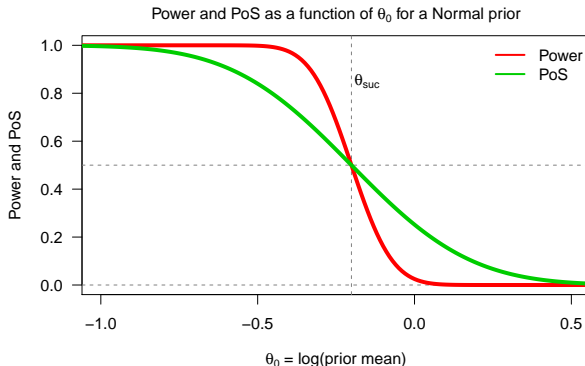
**So what?**

# Question from decision-makers: “PoS is smaller than power?”

Recall example, assuming Phase 2 effect is  $\theta_0 = \log(0.730)$ :

- Power =  $\Phi\left(\frac{\theta_{\text{suc}} - \theta_0}{\sigma_{\text{final}}}\right) = 0.866 \Rightarrow$  PoS with prior = point mass at prior mean.
- PoS =  $\Phi\left(\frac{\theta_{\text{suc}} - \theta_0}{\sqrt{\sigma_{\text{final}}^2 + \sigma_0^2}}\right) = 0.647$ .

$\Rightarrow$  PoS always smaller than power if power  $\geq 0.5$ .



Rufibach et al. (2016a): Extends statement to any **symmetric and unimodal** prior.

# PoS is smaller than power

**4.3. Observation 3: Irrespective of the magnitude of the final sample size, predictive power may not reach desired level (e.g. 80 or 90%)**

Dallow and Fina (2011)

## Bayesian predictive power: choice of prior and some recommendations for its use as probability of success in drug development

Kaspar Rufibach,<sup>\*</sup> Hans Ulrich Burger, and Markus Abt

Bayesian predictive power, the expectation of the power function with respect to a prior distribution for the true underlying effect size, is routinely used in drug development to quantify the probability of success of a clinical trial. Choosing the prior is crucial for the properties and interpretability of Bayesian predictive power. We review recommendations on the choice of prior for Bayesian predictive power and explore its features as a function of the prior. The density of power values induced by a given prior is derived analytically and its shape characterized. We find that for a typical clinical trial scenario, this density has a *u*-shape very similar, but not equal, to a  $\beta$ -distribution. Alternative priors are discussed, and practical recommendations to assess the sensitivity of Bayesian predictive power to its input parameters are provided. Copyright © 2016 John Wiley & Sons, Ltd.

**Keywords:** Bayesian predictive power; conditional power; prior distribution; probability of technical success

**Can we make that work in  
pharma drug development?**

# Lymphoma

- B-cell malignancies:
  - Indolent non-Hodgkin lymphoma (iNHL),
  - diffuse large B-cell lymphoma (DLBCL, aggressive NHL),
  - chronic lymphocytic lymphoma (CLL).
- Standard therapies prior to our new drug Gazyva: rituximab + chemo.
- Accepted primary endpoint: **progression-free survival**.
- Widely believed that response (complete, CR, overall = complete + partial response, ORR) at end of chemo provides indication of efficacy.

# Obinutuzumab, Gazyva, GA101

Obinutuzumab, Gazyva, GA101:

- 2nd generation anti-CD20 antibody.
- Demonstrated single agent activity in iNHL, DLBCL, CLL.
- Targeted to be **best-in-class**: Superior efficacy, supposed to broadly replace rituximab.
- Fast to market - rituximab patent expiration looming. Some **risk appetite**.

GAUSS: Randomized Phase 2 trial in R/R iNHL,

- GA101 mono vs. rituximab mono,
- 149 patients,
- primary endpoint ORR, PFS also collected.

## Four randomized Phase 3 trials:

Trial	Line	Indication	#patients G vs. R	1st patient randomized
<b>CLL11</b>	1st	CLL	356	April 2010
<b>Gadolin</b>	2nd	indolent NHL	413	April 2010
<b>Gallium</b>	1st	indolent NHL	1202	July 2011
<b>Goya</b>	1st	aggressive NHL	1418	July 2011

Primary endpoint for all Phase 3 trials: PFS.

Futility ( $\approx 30\%$  of information) and efficacy interim analysis ( $\approx 67\%$ ) planned for all.

Interim analyses by **trial-specific** independent DMCs.



# Trials

GAUSS: Randomized Phase 2 trial in R/R iNHL,

- GA101 mono vs. rituximab mono,
- 149 patients,
- primary endpoint ORR, PFS also collected.

## Four randomized Phase 3 trials:

Trial	Line	Indication	#patients G vs. R	1st patient randomized
<b>CLL11</b>	1st	CLL	356	April 2010
<b>Gadolin</b>	2nd	indolent NHL	413	April 2010
<b>Gallium</b>	1st	indolent NHL	1202	July 2011
<b>Goya</b>	1st	aggressive NHL	1418	July 2011

Primary endpoint for all Phase 3 trials: PFS.

Futility ( $\approx 30\%$  of information) and efficacy interim analysis ( $\approx 67\%$ ) planned for all.

Interim analyses by **trial-specific** independent DMCs.

## PoS for Goya and Gallium over time

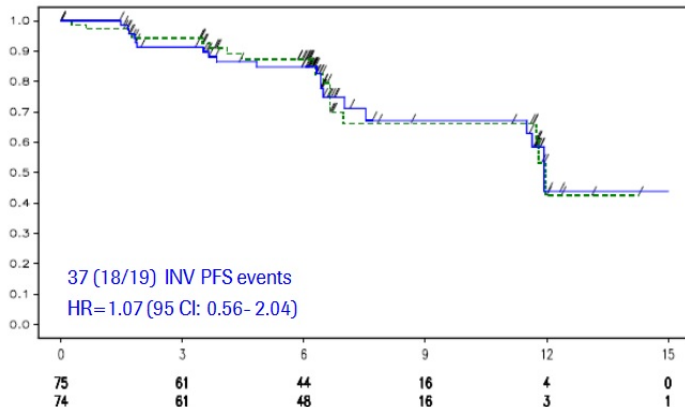
Event	Goya	Gallium
First discussion of trials	0.65	0.65

# Update 1, Q1/2011: GAUSS primary analysis

Goya and Gallium had not started yet.

GAUSS primary analysis:

- Investigator response:  $\Delta$ ORR 4.6%.
- Independently-assessed response:  $\Delta$ ORR 15.2%.
- **PFS hazard ratio 1.07.**



# Update 1, Q1/2011: GAUSS primary analysis

## Actions:

- Based on GAUSS: **decrease PoS qualitatively to 0.41** for both trials.
- Add **early futility interim analyses** to both trials:
  - Endpoints  $\Delta$ CR and  $\Delta$ ORR, **not PFS**.
  - Bars to jump: Goya  $\Delta = 0.05$ , Gallium  $\Delta = 0.03$  for CR proportion in favour of Gazyva.
  - After 200 (Goya) and 170 patients (Gallium) evaluable for response.

## PoS for Goya and Gallium over time

Event	Goya	Gallium
First discussion of trials	0.65	0.65
Update 1, Q1/2011	0.41	0.41

## Update 2, Q1/2013: various results

In Q1/2013, four sources of external knowledge:

- **CLL11**, two final analyses (G vs. standard chemo, R vs. standard chemo).  
 $X_{\text{CLL11}} \sim N(\log 0.44, 4/64.2)$ .
- **GAUSS**, PFS update.  $X_{\text{GAUSS}} \sim N(\log 0.96, 4/76)$ .
- **GALLIUM**, futility interim on CR passed.
- **GOYA**, futility interim on CR passed.

All randomized Gazyva trials, but

- different primary endpoints,
- different phase,
- different indications.
- GAUSS already used with earlier snapshot.

Futility analysis: continue trial if complete response proportion difference among first 170 randomized patients  $\geq 3\%$ .

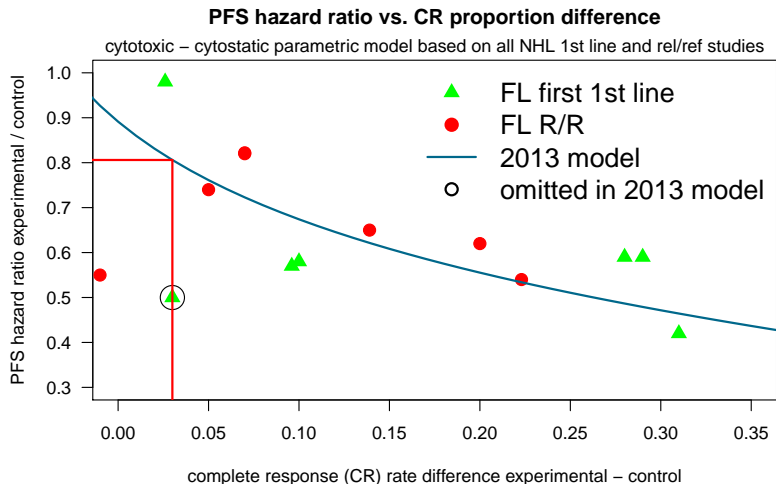
iDMC recommended to continue.

How to include information in PoS computation?

- Parametric model associating  $\Delta$  in CR proportion to PFS based on meta-regression model  $\Rightarrow$  function  $\widehat{HR}(\Delta)$ .
- Use  $\widehat{HR}(\Delta = 3\%)$  as point estimate for hazard ratio.
- Variance: based on 12 PFS events among 170 futility analysis patients.

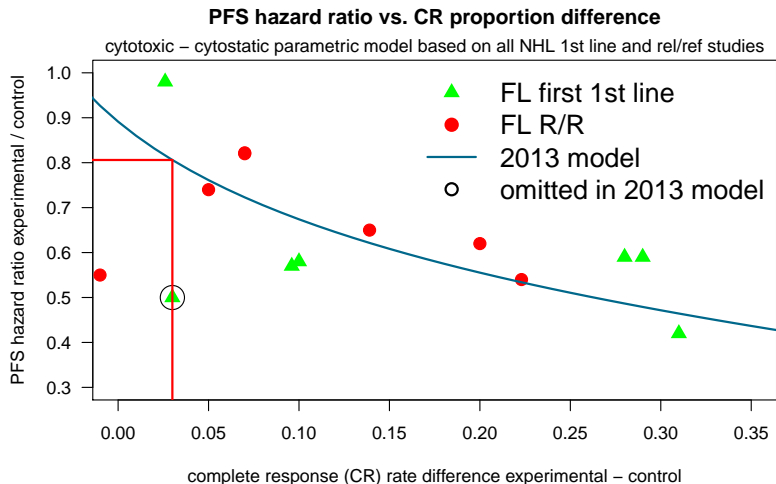
p.s.: iDMC was looking at “totality of information”.

# GALLIUM: NHL 1st line and rel/ref studies



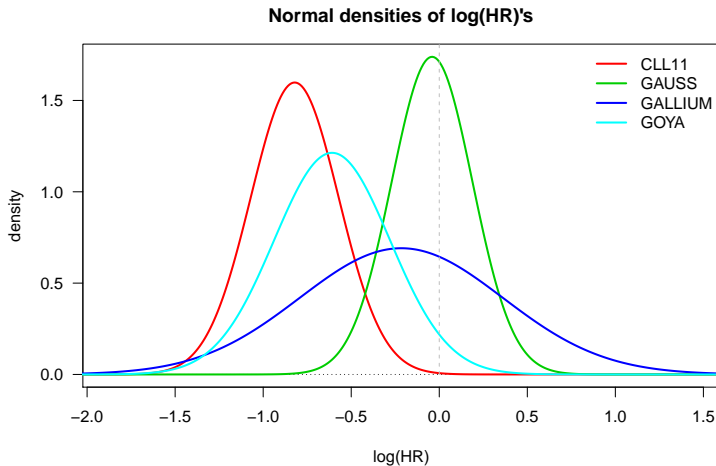


# GALLIUM: NHL 1st line and rel/ref studies



Quantify knowledge from GALLIUM via  
 $X_{\text{GALLIUM}} \sim N(\log 0.806, 4/12).$

# Distributions of $\log(\text{HR})$ 's



We need one data density  $\phi_{\text{data}}$ . Various possibilities to synthesize. See backup.

## Update 2, Q1/2013: Actions

Action senior management based on these results: **Do not modify PoSs.**

- Synthesized information on effect of GA101 vs. rituximab **much depending on chosen weights to combine four results.**
- **Uncertainty** not accounted for: response models, “totality of information” in futilities.
- GAUSS already used to downgrade PoS.
- Nothing had been **pre-specified.**

Action Biostatistics: “Back-engineer” Normal prior:

- Assume initial PoS of 0.41 is worth  $\#PFS$  events observed in futility interim  $\Rightarrow$  determines variance.
- Compute prior mean for each trial from closed formula.
- Yields 0.870 for Gallium, 0.886 for Goya.
- Why different?  $\#events$  at final analysis is different.

## PoS for Goya and Gallium over time

Event	Goya	Gallium
First discussion of trials	0.65	0.65
Update 1, Q1/2011	0.41	0.41
Update 2, Q1/2013	0.41	0.41

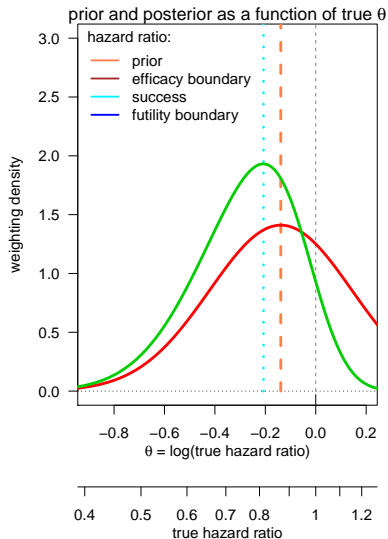
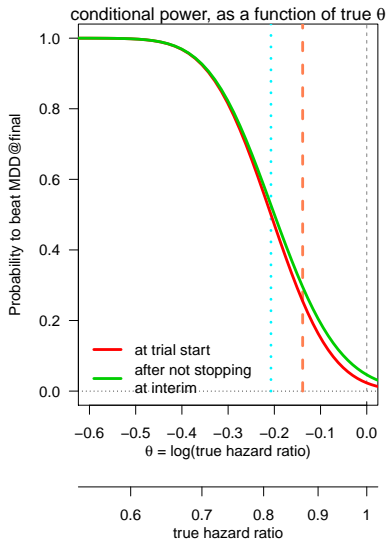
## Update 3: futility interims on PFS

	Goya	Gallium
Number of events	134	111
Futility boundary for hazard ratio	1	1
Date of iDMC decision	21 Mar 2014	28 Jul 2014
iDMC recommendation	<b>Continue</b>	<b>Continue</b>

Action: update PoS with knowledge that **hazard ratio**  $\leq 1$ .

Methodology described in [Rufibach et al. \(2016c\)](#).

# Futility interim - update factors in PoS formula (Gallium)



## PoS for Goya and Gallium over time

Event	Goya	Gallium
First discussion of trials	0.65	0.65
Update 1, Q1/2011	0.41	0.41
Update 2, Q1/2013	0.41	0.41
Update 3, Q1 & 3 2014 (PoS@efficacy interim)	0.49	0.62
Update 3, Q1 & 3 2014 (PoS@final, assume no PFS efficacy)	0.65	0.74
Update 3, Q1 & 3 2014 (PoS@final, assume we pass PFS efficacy)	0.31	0.32

Why PoS lower for Goya?

- More events to define prior  $\Rightarrow$  prior hazard ratio **higher** for Goya.
- Prior density more narrow around this higher prior mean.

## Update 4: efficacy interims on PFS

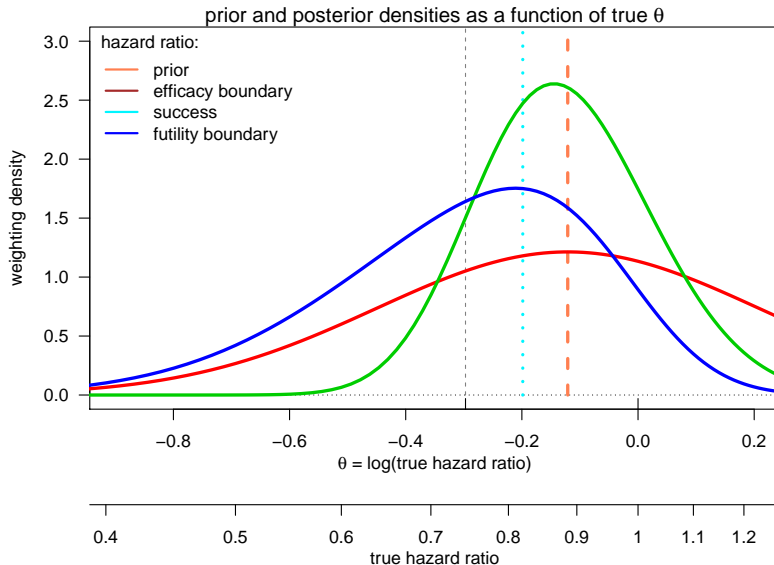
	Goya	Gallium
Number of events	134	248
Efficay boundary for hazard ratio	0.743	0.728
Date of iDMC decision	6 Mar 2015	20 May 2016
iDMC recommendation	<b>Continue</b>	<b>stop and file</b>
Observed hazard ratio		0.66

Action for Goya: update PoS with knowledge that **hazard ratio  $\geq 0.743$** .

Primary publication Gallium: [Marcus et al. \(2017\)](#).



## Efficacy interim - update weighting density (Goya)



## PoS for Goya and Gallium over time

Event	Goya	Gallium
First discussion of trials	0.65	0.65
Update 1, Q1/2011	0.41	0.41
Update 2, Q1/2013	0.41	0.41
Update 3, Q1 & 3 2014 (PoS@efficacy interim)	0.49	0.62
Update 3, Q1 & 3 2014 (PoS@final, assume no PFS efficacy)	0.65	0.74
Update 3, Q1 & 3 2014 (PoS@final, assume we pass PFS efficacy)	0.31	0.32
Update 4, Q1/2015: PoS@final	0.31	

## Update 5: final analysis Goya

	Goya
Number of events	405
Efficacy boundary for hazard ratio	0.820
Unblinding	14 Jul 2016
Observed hazard ratio	<b>0.92</b>

Primary publication Goya: [Vitolo et al. \(2017\)](#).

# Summary of case study

- Example for PoS update throughout lifecycle of drug development program.
- Updates using **external** (other trials) and **internal** (passing of interim analyses) data.
- Quantitative updates **inform** assessment. Stakeholders might not endorse it.  
**Communication** is key!
- Caveats:
  - GAUSS used with two snapshots.
  - Only for PFS interims, we **pre-specified** updates **before** we learned interim decisions.
  - Used point estimates where available, did not account for variability.
  - Uncertainty in response model not accounted for.
  - iDMC issued recommendation taking into "totality of information".
  - Prior "back-engineered", not **formally elicited**.

**What exactly were you  
trying to tell us?**

# Discussion

- Methodology to update PoS after not stopping at an interim analysis.
- $\text{PoS} \neq \text{power} \Rightarrow$  recalibrate stakeholders.
- Density of power values **bathtub-shaped** for typical development scenario.
  - **Sensible** to summarize this distribution in one number which we call BPP?
  - Prior with large variance **not necessarily** uninformative!
- Extension to  $>1$  interims straightforward. Code for two interims in `bpp`.
- Project teams within Roche routinely apply this methodology. Biggest advantage: **systematic assessment** of available evidence.
- No formal prior elicitation at Roche.

# Open questions

All in all nice exercise in applied statistics. Led to:

- Two methodological publications: [Rufibach et al. \(2016a\)](#), [Rufibach et al. \(2016c\)](#).
- R package `bpp`: [Rufibach et al. \(2016b\)](#).

# Open questions

All in all nice exercise in applied statistics. Led to:

- Two methodological publications: [Rufibach et al. \(2016a\)](#), [Rufibach et al. \(2016c\)](#).
- R package `bpp`: [Rufibach et al. \(2016b\)](#).

```
> library(bpp)
> browseVignettes(package = "bpp")
```



# Open questions

All in all nice exercise in applied statistics. Led to:

- Two methodological publications: [Rufibach et al. \(2016a\)](#), [Rufibach et al. \(2016c\)](#).
- R package `bpp`: [Rufibach et al. \(2016b\)](#).

```
> library(bpp)
> browseVignettes(package = "bpp")
```

Open questions:

- How to select prior?
- **Many trials**, i.e. entire company portfolio: How does shape of  $g$  determine fate of your portfolio, or company?
- Explore set of biomarker cutoffs in Phase II, take largest subgroup that still beats TPP. How to reflect this “strategy” in computation of PoS, i.e. how much does PoS need to be penalized? Shrinkage?

**Thank you for your attention.**

# References I

- ▶ Berry, S. M., Carlin, B. P., Lee, J. J. and Müller, P. (2011). *Bayesian adaptive methods for clinical trials*, vol. 38 of *Chapman & Hall/CRC Biostatistics Series*. CRC Press, Boca Raton, FL. With a foreword by David J. Spiegelhalter.
- ▶ Dallow, N. and Fina, P. (2011). The perils with the misuse of predictive power. *Pharm. Stat.* **10** 311–317.
- ▶ Dmitrienko, A. and Wang, M.-D. (2006). Bayesian predictive approach to interim monitoring in clinical trials. *Stat. Med.* **25** 2178–2195.
- ▶ Herson, J. (1979). Predictive probability early termination plans for phase II clinical trials. *Biometrics* **35** 775–783.
- ▶ Jennison, C. and Turnbull, B. W. (2000). *Group sequential methods with applications to clinical trials*. Chapman & Hall/CRC, Boca Raton, FL.
- ▶ Marcus, R., Davies, A., Ando, K., Klapper, W., Opat, S., Owen, C., Phillips, E., Sangha, R., Schlag, R., Seymour, J. F., Townsend, W., Trneny, M., Wenger, M., Fingerle-Rowson, G., Rufibach, K., Moore, T., Herold, M. and Hiddemann, W. (2017). Obinutuzumab for the First-Line Treatment of Follicular Lymphoma. *N. Engl. J. Med.* **377** 1331–1344.
- ▶ O'Hagan, A., Stevens, J. W. and Campbell, M. J. (2005). Assurance in clinical trial design. *Pharm. Stat.* **4** 187–201.
- ▶ Paul, S. M., Mytelka, D. S., Dunwiddie, C. T., Persinger, C. C., Munos, B. H., Lindborg, S. R. and Schacht, A. L. (2010). How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov* **9** 203–214.
- ▶ Proschan, M., Lan, K. and Wittes, J. (2006). *Statistical Monitoring of Clinical Trials: A Unified Approach*. Springer, New York.

# References II

- ▶ Ringel, M., Tollman, P., Hersch, G. and Schulze, U. (2013). Does size matter in R&D productivity? If not, what does? *Nat Rev Drug Discov* **12** 901–902.
- ▶ Rufibach, K., Burger, H. and Abt, M. (2016a). Bayesian predictive power: Choice of prior and some recommendations for its use as probability of success in drug development. *Pharm. Stat.* **15** 438–446.
- ▶ Rufibach, K., Jordan, P. and Abt, M. (2016b). *bpp: Computations Around Bayesian Predictive Power*. R package version 1.0.0.  
<https://CRAN.R-project.org/package=bpp>
- ▶ Rufibach, K., Jordan, P. and Abt, M. (2016c). Sequentially updating the likelihood of success of a Phase 3 pivotal time-to-event trial based on interim analyses or external information. *J Biopharm Stat* **26** 191–201.
- ▶ Spiegelhalter, D., Reedman, L. and Blackburn, P. (1986). Monitoring clinical trials - conditional power or predictive power. *Control Clin Trials* **7** 8–17.
- ▶ Vitolo, U., Trneny, M., Belada, D., Burke, J. M., Carella, A. M., Chua, N., Abrisqueta, P., Demeter, J., Flinn, I., Hong, X., Kim, W. S., Pinto, A., Shi, Y. K., Tatsumi, Y., Oestergaard, M. Z., Wenger, M., Fingerle-Rowson, G., Catalani, O., Nielsen, T., Martelli, M. and Sehn, L. H. (2017). Obinutuzumab or Rituximab Plus Cyclophosphamide, Doxorubicin, Vincristine, and Prednisone in Previously Untreated Diffuse Large B-Cell Lymphoma. *J. Clin. Oncol.* JCO2017733402.

# Backup

# Closed form of PoS if prior is Normal

## Lemma (Explicit computation of PoS)

*Assuming the prior is Normal with density  $q_{\text{prior}}$ , mean  $\theta_0$ , variance  $\sigma_0^2$ , and is independent of the random variable  $\hat{\theta}_{\text{final}}$ . Then*

$$\text{PoS} := \int P_{\theta}(\hat{\theta}_{\text{final}} \leq \theta_{\text{suc}}) q_{\text{prior}}(\theta) d\theta = \Phi\left(\frac{\theta_{\text{suc}} - \theta_0}{\sqrt{\sigma_{\text{final}}^2 + \sigma_0^2}}\right).$$

**Proof:** Use law of total probability and properties of Normal distribution. See [Rufibach et al. \(2016c\)](#).

References containing alternative proofs: [Spiegelhalter et al. \(1986\)](#), [O'Hagan et al. \(2005\)](#), [Proschan et al. \(2006\)](#), or [Dmitrienko and Wang \(2006\)](#).

# Update PoS with external information

Assume we have **external** estimate  $\hat{\theta}_{\text{extern}}$  of treatment effect, with  $\text{SE}(\hat{\theta}_{\text{extern}})$ :

- Study from competitor, collaborative group, ...
- Internal study in same or related program, ...

Quantify knowledge with Normal density  $q_{\text{data}}$ , update prior  $q_{\text{prior}}$  to get  $q_{\text{posterior}}$ .

PoS formula...

$$\text{PoS} = \int_{-\infty}^{\infty} P_{\theta}(\hat{\theta}_{\text{final}} \leq \theta_{\text{suc}}) q_{\text{prior}}(\theta) d\theta$$

...becomes:

$$\text{PoS} = \int_{-\infty}^{\infty} P_{\theta}(\hat{\theta}_{\text{final}} \leq \theta_{\text{suc}}) q_{\text{posterior}}(\theta) d\theta.$$

Simply update prior with external information, recompute PoS.

Power part remains unaffected.

# Update PoS after an unblinded interim analysis

Assume:

- Interim analysis is performed.
- Hazard ratio estimate  $\hat{\theta}_{\text{interim}}$  **known** to the sponsor.

Knowledge **internal** (as opposed to “external”) to the trial!

PoS formula becomes:

$$\text{PoS} = \int_{-\infty}^{\infty} P_{\theta}(\hat{\theta}_{\text{final}} \leq \theta_{\text{suc}} | \hat{\theta}_{\text{interim}} = \theta_{\text{interim}}) q_{\text{prior}}(\theta) d\theta.$$

Power becomes **conditional power**.

Potentially may also want to update  $q_{\text{prior}} \Rightarrow q_{\text{posterior}}$ .

Need to account for **correlation between  $\hat{\theta}_{\text{interim}}$  and  $\hat{\theta}_{\text{final}}$**   $\Rightarrow$  computations based on independent increments property of logrank statistics.

Rufibach et al. (2016c) for details.



# Update PoS after an unblinded interim analysis

Computation:  $\hat{\theta}_{\text{interim}}$  and  $\hat{\theta}_{\text{final}}$  are **correlated**!

By independent increments property of logrank statistic, see [Rufibach et al. \(2016c\)](#):

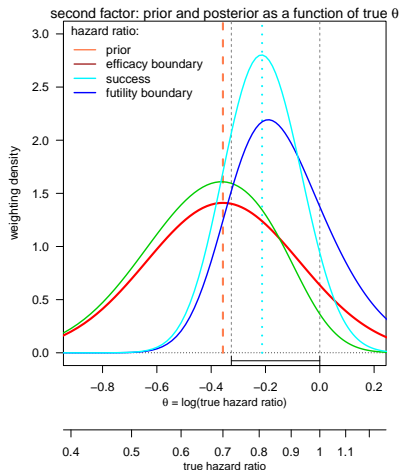
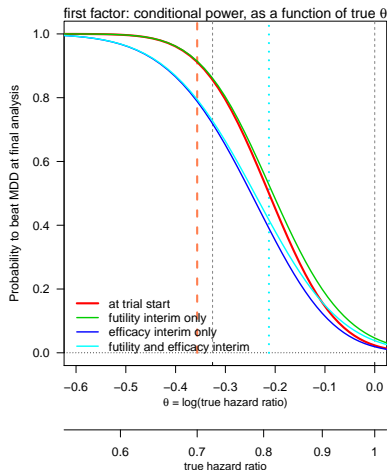
$$\begin{aligned} P_{\theta}(\hat{\theta}_{\text{final}} \leq \theta_{\text{suc}} | \hat{\theta}_{\text{interim}} = \theta_{\text{interim}}) &= \\ &= \Phi\left(\frac{d_{\text{final}}\theta_{\text{suc}} - d_{\text{interim}}\theta_{\text{interim}} - (d_{\text{final}} - d_{\text{interim}})\theta}{\sqrt{4(d_{\text{final}} - d_{\text{interim}})}}\right) \end{aligned}$$

**Conditional power.** [Proschan et al. \(2006\)](#): Derivation based on properties of Brownian Motion (“B-value framework”).

Bayes  $\Rightarrow$  simply plug-in estimate into density:

$$q_{\text{posterior}}(\theta | \hat{\theta}_{\text{interim}} = \theta_{\text{interim}}) \propto \exp\left(-\frac{d_{\text{interim}}}{8}(\theta_{\text{interim}} - \theta)^2\right).$$

# Summarize conditional power and posterior density functions



# Bivariate Normality of $(\hat{\theta}_{\text{interim}}, \hat{\theta}_{\text{final}})$

Page numbers and formula references below refer to [Jennison and Turnbull \(2000\)](#).

After stage  $k$  ( $k = \text{interim}, \text{final}$ ),

- consider logrank teststatistic  $S_k \sim N(\theta d_k/4, d_k/4)$  (formula 3.15, p. 78ff),
- and look at the transformation  $\hat{\theta}_k = S_k/(d_k/4) \sim N(\theta, 4/d_k)$  (= log hazard ratio estimate, p. 78/79).

Apply **standard group sequential** framework for Normal teststatistics to get bivariate Normality of  $(\hat{\theta}_{\text{interim}}, \hat{\theta}_{\text{final}})$  and covariance between  $\hat{\theta}_{\text{interim}}$  and  $\hat{\theta}_{\text{final}}$ .

Covariance: Write  $\hat{\theta}_k$  as sum of Normals, compute covariance explicitly (p. 49/50).

Alternatively:

- Use general **conditional power** framework, as outlined in [Proschan et al. \(2006\)](#) (see p. 20 for an explicit computation).
- Not clear how to adapt to “interval knowledge only”, though.

# Continuous and binary endpoints

## Continuous endpoint:

- Known variance  $\Rightarrow$  setup dealt with above.
- Unknown variance: assess sensitivity to various assumed variances.

## Binary endpoint, see discussion in O'Hagan et al. (2005):

- Effect estimate also depends on proportion  $\pi_1$  in control group  $\Rightarrow$  sensitivity analysis w.r.t. to  $\pi_1$ .
- Power function can only be derived if Normal approximation is assumed to be exact (i.e. proportions plugged-in in variance are assumed known)  $\Rightarrow$  **uncertainty in variance estimation ignored.**
- No closed formula for power. Numerically compute PoS using joint distribution of proportions in control and treatment group.
- Want to account for all uncertainties, i.e. also in variance estimation  $\Rightarrow$  simulate.
- Choice of prior? [Herson \(1979\)](#): "Uniform priors give too much weight on extreme cases"  $\Rightarrow$  bathtub!

# Update after more than one interim analysis

What does change?

$$\begin{pmatrix} \hat{\theta}_{\text{interim}} \\ \hat{\theta}_{\text{final}} \end{pmatrix} \sim N\left(\begin{pmatrix} \theta \\ \theta \end{pmatrix}, \begin{pmatrix} 4/d_{\text{interim}} & 4/d_{\text{final}} \\ 4/d_{\text{final}} & 4/d_{\text{final}} \end{pmatrix}\right)$$

becomes

$$\begin{pmatrix} \hat{\theta}_{\text{int1}} \\ \hat{\theta}_{\text{int2}} \\ \hat{\theta}_{\text{final}} \end{pmatrix} \sim N\left(\begin{pmatrix} \theta \\ \theta \\ \theta \end{pmatrix}, \begin{pmatrix} 4/d_{\text{int1}} & 4/d_{\text{int2}} & 4/d_{\text{final}} \\ 4/d_{\text{int2}} & 4/d_{\text{int2}} & 4/d_{\text{final}} \\ 4/d_{\text{final}} & 4/d_{\text{final}} & 4/d_{\text{final}} \end{pmatrix}\right).$$

**Blinded** interim: Compute conditional power  $P_{\theta}(\hat{\theta}_{\text{final}} \leq \theta_{\text{suc}} | \hat{\theta}_1 \in \mathcal{I}_1, \hat{\theta}_2 \in \mathcal{I}_2)$  and updated weighting density  $q_1(\theta | \hat{\theta}_1 \in \mathcal{I}_1, \hat{\theta}_2 \in \mathcal{I}_2)$ .

Code available.

**Unblinded** interim: Not worked out yet.

# Interim analysis based on surrogate endpoint

## MIRROS study:

- Acute myeloid lymphoma (AML), 2nd line.
- Phase 3, 2:1 randomized, compare Idasanutlin + CT vs. CT alone.
- Primary endpoint **OS**.
- Final analysis planned after **275 events** to detect HR 0.667 with 85% power.

## Special features of MIRROS:

- Power computation takes into account **cure proportions** in both arms.
- Futility interim analysis planned after **120 patients** (LIP enabling).
- Interim passed if either (OR for CR  $\geq 2.5$ ) or (HR for EFS  $\leq 1$  and OR for CR  $\geq 2$ ).

## Simulation models:

- Sample size received via simulation  $\Rightarrow$  to account for cure proportion.
- Stopping probabilities at interim received via simulations in mechanistic model associating **CR  $\Rightarrow$  EFS  $\Rightarrow$  OS**.

# Interim analysis based on surrogate endpoint

Initial PoS:

- Product of qualitative reference value for small molecule cPoS for Phase 2 (35%) and reference values for cPoS for Phase 3 (65%)  $\Rightarrow$  **0.2275**.
- Tune “pessimistic prior” from [Rufibach et al. \(2016c\)](#) to get initial PoS to beat TPP equal to 0.2275.

Question: How to update initial PoS (OS) if we pass interim (CR, EFS)?

- 1 Take simulation model.
- 2 Choose parameters that reflect the **alternative hypothesis** used for sample size planning and simulate 10'000 trials.
- 3 Look at distribution of OS HRs for those simulated scenarios that jump the interim hurdle based on CR and EFS.
- 4 90% of these OS HRs are  $\leq 0.993$ , 80% are  $\leq 0.865$ .
- 5 Report BPP update for these boundaries.

# Efficacy interim analysis only

At efficacy interim not stopped with boundary  $HR > 0.722$ : we know that

- $0.722 \leq HR$  or
- $\hat{\theta}_{\text{interim}} \in (\log(0.722), \infty)$ .

How does PoS change after this interim?



# Efficacy interim analysis only

At efficacy interim not stopped with boundary  $HR > 0.722$ : we know that

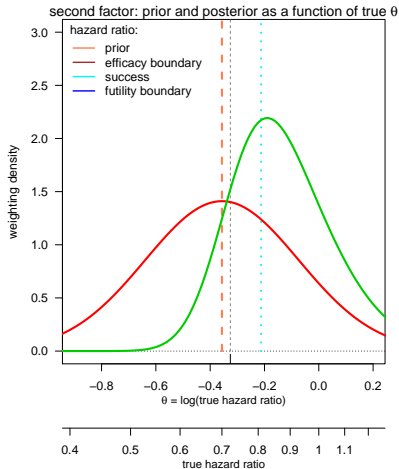
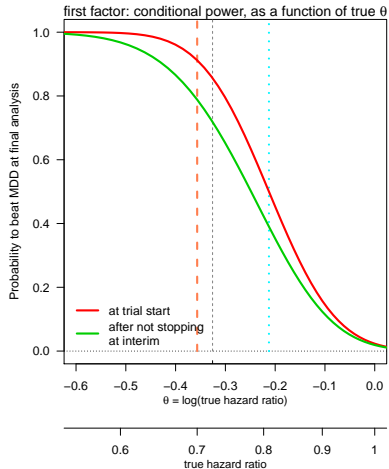
- $0.722 \leq HR$  or
- $\hat{\theta}_{\text{interim}} \in (\log(0.722), \infty)$ .

How does PoS change after this interim?

PoS decreases from **0.683** to **0.317**.

Why?

# Efficacy interim analysis only - plot both factors in PoS formula



- Green density **not** a Normal density, but product of difference of values of Normal CDF and prior density  $q_{\text{prior}}$ .
- Centering of posterior at  $\approx \theta_{\text{suc}}$  is **coincidental!**

# Efficacy interim analysis only - sensitivity analysis

PoS at trial start: 0.683.

In our example, with  $\theta_{\text{efficacy}} = \log(0.722)$  and  $\theta_{\text{futility}} = \infty$ :

$\hat{\theta}_{\text{interim}}$	value of $\hat{\theta}_{\text{interim}}$	PoS after not stopping
$\in (\theta_{\text{efficacy}}, \theta_{\text{futility}})$	$(\log(0.722), \infty)$	0.317
$= \theta_{\text{efficacy}}$	$\log(0.722)$	0.944

# Efficacy interim analysis only - comments

After not stopping at interim, PoS decreases from **0.683** to **0.317**.

Why does PoS decrease after not stopping?

- Prior with prior mean  $\log(0.7)$  assigns weight to hazard ratios smaller than hazard ratio to finally beat ( $\theta_{\text{suc}} = \log(0.809)$ ).
- **Not stopping** shifts mass of prior  $q_{\text{prior}}$  to the right of 0.722 for  $q_{\text{posterior}} \Rightarrow$  more weight on hazard ratios  $> \theta_{\text{suc}}$ .
- Together with decrease in conditional power accounts for **lower PoS** after not stopping.
- Amount of decrease largely depends on information fraction.
- Rule-of-thumb: Expect PoS around 1/3 if not stopping after efficacy interim after 2/3 of information.

# Futility and efficacy interim analysis

Interim passed with boundaries  $0.722 < HR \leq 1$ : we know that

- $0.722 < HR \leq 1$  or
- $\hat{\theta}_{\text{interim}} \in (\log(0.722), \log(1)]$ .

How does PoS change after this interim?

# Futility and efficacy interim analysis

Interim passed with boundaries  $0.722 < \text{HR} \leq 1$ : we know that

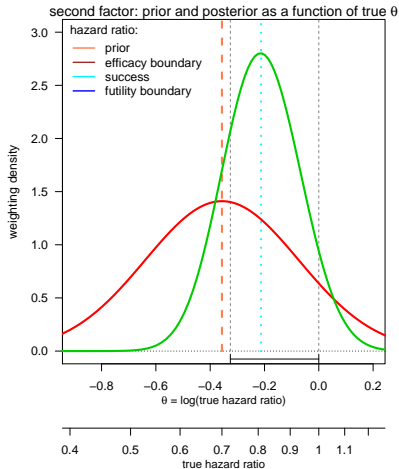
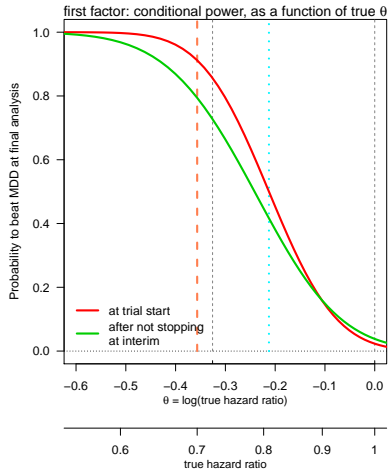
- $0.722 < \text{HR} \leq 1$  or
- $\hat{\theta}_{\text{interim}} \in (\log(0.722), \log(1)]$ .

How does PoS change after this interim?

PoS decreases from **0.683** to **0.437**.

Why?

# Futility and efficacy interim analysis - factors in PoS formula



- Green density **not** a Normal density, but product of difference of values of Normal CDF and prior density  $q_{\text{prior}}$ .
- Centering of posterior at  $\approx \theta_{\text{suc}}$  is **coincidental!**

# Futility and efficacy interim analysis - sensitivity analysis

PoS at trial start: 0.683.

In our example, with  $\theta_{\text{efficacy}} = \log(0.722)$  and  $\theta_{\text{futility}} = \log(1)$ :

$\hat{\theta}_{\text{interim}}$	value of $\hat{\theta}_{\text{interim}}$	PoS after not stopping
$\in (\theta_{\text{efficacy}}, \theta_{\text{futility}}]$	$(\log(0.722), \log(1)]$	0.437
$= \theta_{\text{efficacy}}$	$\log(0.722)$	0.944
$= (\theta_{\text{efficacy}} + \theta_{\text{futility}})/2$	$\log(0.850)$	0.298
$= \theta_{\text{futility}}$	$\log(1)$	0.004



# Futility and efficacy interim analysis - comments

After not stopping at interim, PoS decreases from **0.683** to **0.437**.

Why does PoS decrease after not stopping?

- Prior with prior mean  $\log(0.700)$  assigns weight to hazard ratios smaller than hazard ratio to finally beat ( $\theta_{\text{suc}} = \log(0.809)$ ).
- **Not stopping** shifts mass of prior  $q_{\text{prior}}$  to the right of 0.722 and left of 1.000 for  $q_{\text{posterior}} \Rightarrow$  posterior “squeezed” between boundaries, in this case  $\approx$  centered at  $\theta_{\text{suc}}$ .
- Bit more mass to the right of  $\theta_{\text{suc}}$ . Together with small decrease in conditional power accounts for **lower PoS** after not stopping.

Phase III randomized **3-arm** trial in CLL.

Arms: Chlorambucil (C), rituximab + C, Gazyva + C, randomization 1:2:2.

Two results, based on **different** snapshots (Stage 1a, Stage 1b):

- Stage 1a:  $\hat{\lambda}_{GvsC} = 0.14$ , based on **123** events.  
 $SE(\log \hat{\lambda}_{GvsC}) = 2.12 \cdot 123^{-1/2} = 0.19$ .
- Stage 1b:  $\hat{\lambda}_{RvsC} = 0.32$ , based on **175** events.  
 $SE(\log \hat{\lambda}_{RvsC}) = 2.12 \cdot 175^{-1/2} = 0.16$ .
- Assuming **Exponentiality**:  $\hat{\lambda}_{GvsR} \approx \hat{\lambda}_{GvsC} / \hat{\lambda}_{RvsC} = 0.44$ .

Standard error of  $\log \hat{\lambda}_{GvsR} = \log \hat{\lambda}_{GvsC} - \log \hat{\lambda}_{RvsC}$ :

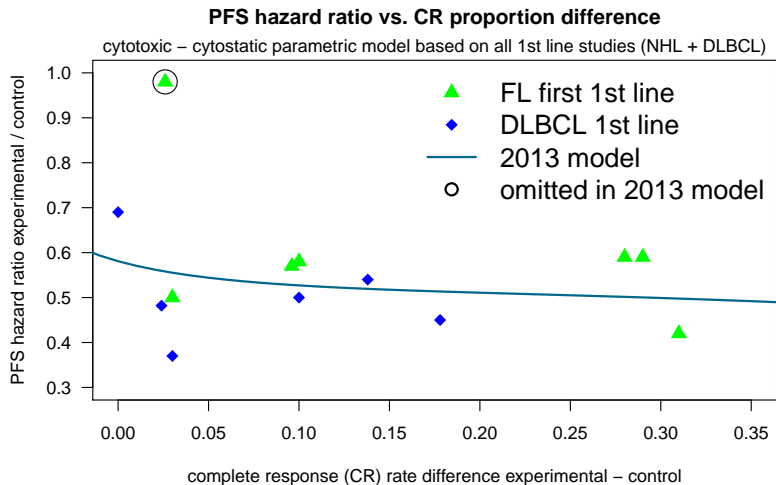
$$\begin{aligned}
 SE(\log \hat{\lambda}_{GvsR}) &= \sqrt{SE(\log \hat{\lambda}_{GvsC})^2 + SE(\log \hat{\lambda}_{RvsC})^2} \\
 &= \sqrt{0.19^2 + 0.16^2} \\
 &= 0.25 = \sqrt{4/64.2}.
 \end{aligned}$$

Quantify knowledge:  $\mathbf{X}_{CLL11} \sim \mathbf{N}(\log 0.44, 4/64.2)$ .

Updated PFS analysis:

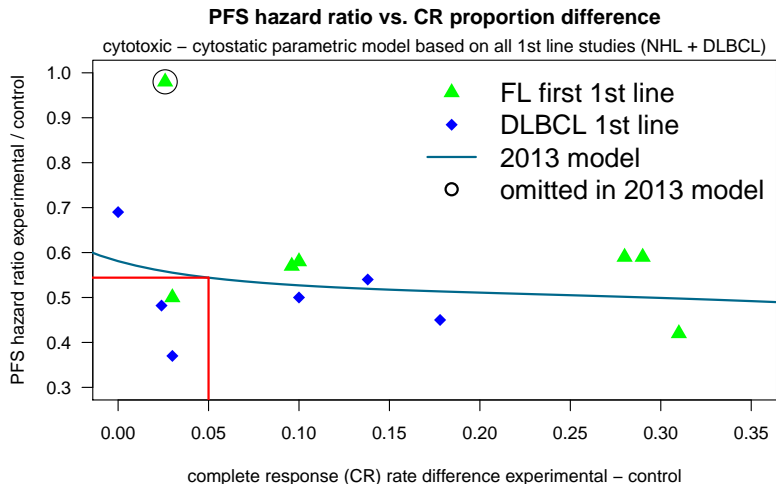
$$\mathbf{X}_{\text{GAUSS}} \sim \mathbf{N}(\log 0.96, 4/76).$$

# GOYA: NHL and DLBCL 1st line studies



- Futility analysis: continue trial if complete response proportion difference among first 200 randomized patients  $\geq 5\%$ .
- Same approach as GALLIUM.

# GOYA: NHL and DLBCL 1st line studies



Quantify knowledge from GOYA via  
 $X_{\text{GOYA}} \sim N(\log 0.544, 4/37)$ .

# Synthesizing different sources

## Weighted mean:

- different indications,
- first line vs. relapsed/refractory,
- GOYA & GALLIUM estimates based on model associating HR to CR rate difference, based on **few** studies only.

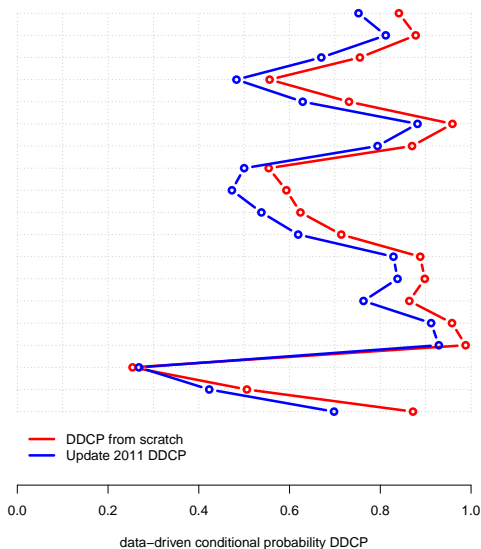
Weighted mean assigns each study the importance one believes it has for GOYA / GALLIUM.

Various approaches: elicit weights from clinicians, meta-analysis, extreme scenarios, ...

# Discussed scenarios

DDCP for GALLIUM at the end of February 2013, depending on weight scenario

- 1: all equal
- 2: GOYA and GALLIUM low
- 3: NHL high 1
- 4: NHL high 2
- 5: omit GOYA, others equal
- 6: omit CR futilites, CLL11 high
- 7: omit CR futilites, others equal
- 8: omit CR futilites, GAUSS high
- 9: only consider futilities
- 10: omit CLL11, others equal
- 11: CLL11 10percent, others equal
- 12: Meta analysis fixed effects
- 13: Meta analysis random effects
- 14: Michael Wenger
- 15: Nancy Valente
- 16: CLL11 only
- 17: GAUSS only
- 18: GALLIUM modelling only
- 19: GOYA modelling only



**Backup slides: parametric cytotoxic-cytostatic model.**



# Quantities

E: experimental arm, S: standard arm.

Response proportions:

- $\theta_S$ : Probability of response in standard treatment.
- $\theta_E$ : Probability of response in experimental treatment.

Quantity that inputs the model (the “x” in  $f(x)$ ):

$$\Delta = \theta_E - \theta_S.$$

Further derived quantities:

- $\lambda$ : Hazard in standard treatment, non-responders.
- $c$ : Hazard ratio experimental vs. standard treatment.
- $r$ : Hazard ratio responders vs. non-responders.

# Statistical model

We assume **proportional hazards** for E vs. S and responders vs. non-responders:

- Hazard in experimental treatment, non-responders:  $\lambda_{E, \text{non-resp}} := c \cdot \lambda$ .
- Hazard in standard treatment, responders:  $\lambda_{S, \text{resp}} := r \cdot \lambda$ .
- Hazard in experimental treatment, responders:  $\lambda_{E, \text{resp}} := c \cdot r \cdot \lambda$ .

Hazard ratio for experimental vs. standard:

$$\begin{aligned} \text{HR}_{E \text{ vs. } S} &= \frac{\lambda_E}{\lambda_S} = \frac{c\lambda(1 - \theta_E) + cr\lambda\theta_E}{\lambda(1 - \theta_S) + r\lambda\theta_S} \\ &= c \left( 1 + \frac{\Delta(r - 1)}{(r - 1)^{-1} + \theta_S} \right). \end{aligned}$$

Function  $\text{HR}_{E \text{ vs. } S}(\Delta)$  that relates  $\text{HR}_{E \text{ vs. } S}$  to response proportion difference  $\Delta$ .

# Statistical model

Hazard ratio  $HR_{E \text{ vs. } S}$  as a function of response proportion difference  $\Delta$ :

$$HR_{E \text{ vs. } S}(\Delta) = c \left( 1 + \frac{\Delta(r-1)}{(r-1)^{-1} + \theta_S} \right).$$

Quantities:

- $\theta_S$ : Response proportion standard treatment, assumed based on **historical data**.
- $r$ : Hazard ratio responders vs. non-responders. Will be estimated based on model (see below).
- $c$ : Let  $c_0$  be basic cytostatic effect, typically assumed from historical data.
  - Carreras et al (2011): Assumed  $c = c_0(1 - p\Delta)$ , with  $p$  link factor.
  - GAZYVA gating in 2011:  $c = c_0 \cdot OR_{E \text{ vs. } S}^p$ , with  $OR_{E \text{ vs. } S}$  odds ratio E vs. S,  $p$  link factor. **"Better fit"** for small values of  $\Delta$ .

# GAZYVA specific model

Model used in **GAZYVA 2011 gating**:

$$\text{HR}_{\text{E vs. S}}(\Delta) = c_0 \cdot \text{OR}_{\text{E vs. S}}^p \left( 1 + \frac{\Delta(r-1)}{(r-1)^{-1} + \theta_S} \right).$$

Parameters assumed from historical data:

- $\theta_S$ : response proportion in standard treatment.

Input data:

- $\Delta$ : response proportion difference for each study. The  $x_i$ 's in the regression model.
- $\text{HR}_{\text{E vs. S}}$ : hazard ratio for each study. The  $y_i$ 's in the regression model.
- $\text{OR}_{\text{E vs. S}}$ : additional study-specific input data.

Parameters to be estimated:  $c_0, p, r$ .

# GAZYVA specific model: estimation of parameters

Estimation of parameters  $r$ ,  $c_0$ ,  $p$  via **non-linear regression model**:

- General model as in textbooks:

$$y_i = f(\beta, \mathbf{x}'_i) + \varepsilon_i.$$

- 2011: `proc nlmixed` in SAS. 2013: Ported to R.
- No random effects, so **not a mixed** but simple non-linear regression model.
- $\varepsilon_i$ : independent Normal with mean 0 and variance  $\sigma^2$ .

Specified to GAZYVA gating setup:

$$\text{HR}_{\text{E vs. S}, i}(\Delta) = c_0 \cdot \text{OR}_{\text{E vs. S}, i}^p \left( 1 + \frac{\Delta_i(r-1)}{(r-1)^{-1} + \theta_S} \right).$$

“Observations” in regression model are **studies**, so  $i$  runs through studies under consideration.

Response model: studies not **weighted** according to size.

**Backup slides: synthesizing evidence for Update 2.**

# Synthesizing different sources

Compute **weighted mean** of different sources:

$$X_{\text{data}} = w_{\text{CLL11}}X_{\text{CLL11}} + w_{\text{GAUSS}}X_{\text{GAUSS}} + \\ + w_{\text{GALLIUM}}X_{\text{GALLIUM}} + w_{\text{GOYA}}X_{\text{GOYA}},$$

weights sum up to 1.

Assuming these random variables are **independent** (...)

$\Rightarrow X_{\text{data}}$  Normal with mean

$$\mathbb{E}(X_{\text{data}}) = w_{\text{CLL11}} \log(0.44) + w_{\text{GAUSS}} \log(0.96) + \\ + w_{\text{GALLIUM}} \log(0.806) + w_{\text{GOYA}} \log(0.544)$$

and variance

$$\text{Var}(X_{\text{data}}) = w_{\text{CLL11}}^2(4/64.2) + w_{\text{GAUSS}}^2(4/76) + \\ w_{\text{GALLIUM}}^2(4/12) + w_{\text{GOYA}}^2(4/37).$$

Denote density of  $X_{\text{data}}$  by  $\phi_{\text{data}}$ .

# Finally compute PoS

Now have all ingredients to compute PoS for **GALLIUM**:

- $Y_{\text{final}} \sim N(\theta, \sigma_{\text{final}}^2 = 4/n_{\text{final}})$ , i.e.  $n_{\text{final}}$ : Protocol. ✓
- $\lambda_{\text{MD}}$ : Protocol, target product profile. ✓
- $\phi_{\text{data}}$ .



# Finally compute PoS

Now have all ingredients to compute PoS for **GALLIUM**:

- $Y_{\text{final}} \sim N(\theta, \sigma_{\text{final}}^2 = 4/n_{\text{final}})$ , i.e.  $n_{\text{final}}$ : Protocol. ✓
- $\lambda_{\text{MD}}$ : Protocol, target product profile. ✓
- $\phi_{\text{data}}$ . ✓

Two approaches to finally compute PoS:

- 1 Simply use  $\phi_{\text{data}} = \phi_{\text{data}}$ . Compute PoS from scratch, **no update**.

$$\text{DDCP} = \int_{-\infty}^{\infty} P_{\theta}(Y_{\text{final}} \leq \log \lambda_{\text{MD}}) \phi_{\text{data}}(\theta) d\theta.$$

- 2 Quantify initial 2011 PoS (= 0.41) via  $N(\lambda_{\text{prior}}, \sigma_{\text{prior}}^2)$  with density  $\phi_{\text{prior}}$ . Update this prior with  $\phi_{\text{data}}$  in a **conjugate Normal Bayesian model** to get  $\phi_{\text{data}} = \phi_{\text{posterior}}$ .

$$\text{DDCP} = \int_{-\infty}^{\infty} P_{\theta}(Y_{\text{final}} \leq \log \lambda_{\text{MD}}) \phi_{\text{posterior}}(\theta) d\theta.$$

# *Doing now what patients need next*

**R version and packages used to generate these slides:**

R version: R version 3.5.1 (2018-07-02)

Base packages: stats / graphics / grDevices / utils / datasets / methods / base

Other packages: bpp / mvtnorm / reporttools / xtable

This document was generated on 2019-05-21 at 15:27:17.